

MPEG-7和 MPEG-7实验模型参考软件*)

单锦来 陈博 杨献春 许满武

(南京大学软件新技术国家重点实验室 南京大学计算机科学与技术系 南京210093)

MPEG-7 and MPEG-7 eXperimentation Model Reference Software

SHAN Jin-Lai CHEN Bo YANG Xian-Chun XU Man-Wu

(State Key Lab for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing 210093)

Abstract MPEG-7 is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group), MPEG-7, formally named "Multimedia Content Description Interface", aims to create a standard for describing the multimedia content data. In this paper, firstly we introduce the context of MPEG-7, MPEG-7 objectives and parts. Then we analyze the MPEG-7 eXperimentation Model (XM) reference software, including application types, software models and the implementation of descriptors and description schemes. Finally we discuss the possible application of MPEG-7 in AVTMS.

Keywords MPEG-7, XM reference software, Descriptor, Description scheme

1 MPEG-7简介

1.1 产生背景

随着多媒体技术和网络技术的飞速发展,越来越多的音频、视频信息成为可用。这些信息以诸如静态图片、图形、3D模型、音频和视频等不同格式存在。在过去,人们是直接使用音、视频信息的,如今由计算机创建、交换、检索和使用的音、视频信息增长得非常迅速。音、视频信息在我们的生活中扮演着一个越来越重要的角色,对这些信息进行处理的需求也越来越多。简单的波形方式、基于样本、基于压缩(如 MPEG-1和 MPEG-2)甚至基于对象(MPEG-4)的表示形式已不能满足需求,有必要开发出新的信息表示形式。新的表示形式还要能够对所表示的信息给予一定程度的解释。这种格式的信息能够被一个设备或者一段计算机代码访问。

1.2 目标

MPEG-7就是支持这些处理需求的描述多媒体内容数据的标准。MPEG-7仅标准化对多媒体内容的描述,不标准化使用这些多媒体内容的应用。MPEG-7不是具体针对某个方面的单个应用而是针对尽可能多的应用。

MPEG-7的目标是标准化多媒体环境下描述音视频信息的核心技术。并将通过包括更多的数据类型来扩展现有的标识内容的所有者解决方案的有限的的能力。MPEG-7描述的音视频数据包括静态图片,图形,3D模型,演讲,音频,视频,以及关于这些元素如何组合在多媒体表示中的合成信息。MPEG-7描述工具不依赖于描述内容存储或编码的方式。MPEG-7标准是建立在其他表示法如 analogue, PCM, MPEG-1, -2和-4基础上的。MPEG-7标准的一个功能是提供对这些表示法的相应部分的引用,如 MPEG-4中的一个形状描述符也可以在 MPEG-7环境下使用。MPEG-7允许不同的描述粒度,可以提供不同级别的辨别能力。根据应用领域的不同,

相同的材料可以使用不同类型的特征描述。对于视频信息来说,较低的抽象级别是对形状,大小,质地,颜色,移动和位置的描述。对音频而言包括声音空间中的音调,语气,节奏,节奏的改变和位置等的描述。高层次的描述给出的是语义方面的信息,如“在一座大厦的后面停着一辆白色的小汽车,小汽车前部坐着两个人,他们正在大声说话”。所有的这些描述都被有效地--对搜索有效地进行编码。

低级别的特征可以完全自动地被抽取,一些高级别的特征的抽取需要人机交互才能完成。

1.3 组成

MPEG-7详细规定了一套标准的描述符(Descriptors)来描述各种的多媒体信息。并且预先定义了描述子的结构以及它们之间的关系。这种结构叫做描述模式(Description Schemes),并通过描述定义语言(Description Definition Language DDL)来定义新的描述模式。

MPEG-7标准化了以下元素:

- Descriptors (Ds) 描述符:各种特征的表示,定义每个特征表示的语法和语义。

- Description Schemes (DSs) 描述模式,详细说明 Ds 和 DSs 之间关系的结构和语义。

- Description Definition Language (DDL) 描述定义语言,用于创建新的 DSs 和新的 Ds,以及允许对存在的 DSs 进行扩展和修改。

- System tools 系统工具,支持描述的复合,同步问题,传输机制,文件格式等。

MPEG-7-- Multimedia Content Description Interface 标准包括下面几个部分:

系统 MPEG-7 Systems 包括了一些工具,这些工具用来准备有效地传输和存储 MPEG-7描述,确保内容和描述之间的同步,管理和保护知识产权。它定义终端结构和标准化接

*) 本文受国家自然科学基金(批准号为69973020)和国家863计划课题(课题编号为2001AA111281-2)资助。单锦来 研究生,主要研究方向为新型程序设计。陈博 研究生,主要研究方向为新型程序设计。杨献春 副教授,主要研究方向为新型程序设计。许满武 教授,博士生导师,主要研究方向为新型程序设计。

口。

描述定义语言(DDL) DDL 是 MPEG-7 标准的核心部分。它提供了坚实的描述基础,用户可以通过它来定义自己的 DSs 和 Ds。DDL 定义表达和组合 DSs 和 Ds 的语法规则。

XML 模式语言被 MPEG 选为 DDL 的基础,并在其基础上作了一些扩展,DDL 可以分成下面几个逻辑上标准化的部分:

- The XML Schema structural language components, XML Schema 结构语言成分描述 XML 文档的结构和应遵从的规范。

- The XML Schema datatype language components, XML Schema 数据类型语言成分描述 XML 文档中能够使用的数据类型。

- The MPEG-7 specific extensions, MPEG-7 特定的扩展,描述为了满足 MPEG-7 的需要而扩展的数据类型。

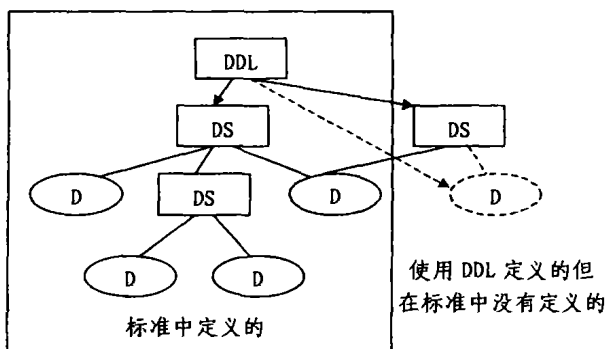


图1 MPEG-7中不同元素之间的关系

视觉 定义描述视觉数据信息的描述符和描述模式 (Ds 和 DSs)。

MPEG-7 Visual 描述工具包括如下方面的基本结构和描述符:1. Color 颜色;2. Texture 质地;3. Shape 形状;4. Motion 动作;5. Localization 位置;6. Others 其他。每一类包括基本的和复杂的描述符。

音频 定义描述音频数据信息的描述符和描述模式 (Ds 和 DSs)。

MPEG-7 Audio 描述工具包括了六种技术:

1. 音频描述框架,2. 声效描述工具,3. 乐器音质描述工具,4. 口语内容描述工具,5. 统一的静默片段,6. 便于 query-by-humming 的旋律描述符。

多媒体描述模式(MDS) MDS 标准化了一套描述工具来处理一般性的多媒体实体。一般性实体是用于视频、音频和文本描述中的特征,适用于所有媒体。

除了一般性媒体描述工具外,更多复杂的描述工具也被标准化了,它们被用来描述多于一种媒体的情况。这些描述工具根据它们的功能可分为5类:

1. Content description 内容描述:表示可察觉的信息。
2. Content management 内容管理:关于媒体特征的信息,AV 特征的创建和使用。
3. Content organization 内容组织:几种 AV 内容的分析和分类的表示。
4. Navigation and access 导航和访问:AV 内容摘要和变化的规范。

5. User interaction 用户交互:描述与多媒体信息消耗相关的用户偏好和使用历史。

参考软件 MPEG-7 相关部分的软件实现。

eXperimentation Model (XM) 软件是 MPEG-7 描述子 (Ds)、描述模式 (DSs)、编码模式 (CSs) 和描述定义语言 (DDL) 的模拟平台,除了标准化的成分,模拟平台需要一些非标准化的成分在数据结构上执行一些过程式的代码。这些数据结构和过程式代码组成了应用。XM 应用分成服务器应用和客户端应用。

Conformance 为测试 MPEG-7 实现的遵从性指定指导方法和过程。

抽取和使用 为描述信息的抽取和使用提供指导和例子。

2 MPEG-7参考软件

参考软件,又叫做 experimentation software (实验软件),是 MPEG-7 标准参考代码的基础框架,它实现标准的 MPEG-7 组件,包括描述符 (Descriptors Ds),描述模式 (Description Schemes DSs),编码模式 (Coding Schemes CSs),描述定义语言 (DDL) 和 BiM 系统组件。根据实现的对象不同可以分成四类:

1. DDL 剖析器和 DLL 验证剖析器 (DDL parser and DDL validation parser)

2. 视觉描述符 (Visual descriptors) 这个模块实现所有的视觉描述符。从视觉媒体中创建视觉描述符,用于抽取视觉描述符的技术是非标准化的,这些抽取工具的质量和复杂度没有进行优化。

3. 音频描述符 (Audio descriptors) 这个模块实现所有的音频描述符。从音频媒体中创建音频描述符,用于音频抽取描述符的技术是非标准化的,这些抽取工具的质量和复杂度没有进行优化。

4. 多媒体描述模式 (Multimedia Description Schemes) 提供 MPEG-7 中所有多媒体描述模式的标准描述。

2.1 MPEG-7参考软件应用类型

MPEG-7 参考软件主要包括以下几种应用类型:

1. 抽取应用 以原始的多媒体数据为处理对象,使用相应媒体的解码器将多媒体数据解码成能够直接处理的中间格式,如将 MPEG-1 的视频信息解码成帧格式。调用相应的描述符或者描述模式的抽取工具对这些中间格式的媒体数据进行特征抽取,填充到对应描述符或描述模式中,并按照指定的编码模式以文件形式存储起来,组成描述数据库。

其过程如图2所示:

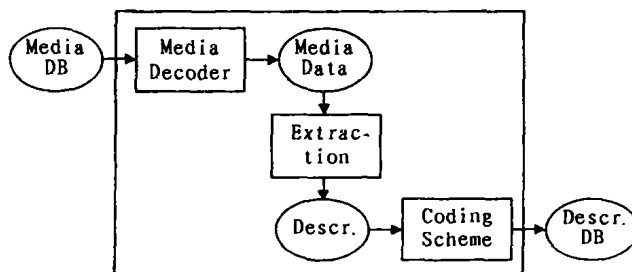


图2

2. 搜索和获取应用 以抽取应用生成的描述数据库作为处理对象。将各种描述信息按照相应的编码模式从文件中读取出来填充到相应的描述符或描述模式对象中去, 经由相应描述符或描述模式的匹配工具与用户查询输入所生成的描述符或描述模式对象进行比较匹配, 最终返回一个最为匹配的结果列表。通过这个列表用户可以获取到相应的媒体。

其过程如图3所示:

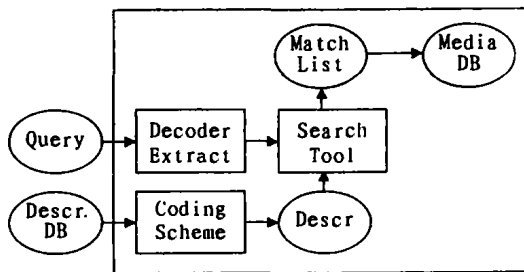


图3

3. 转换代码应用 将媒体数据库和媒体描述数据库加载到内存中, 将媒体描述数据与输入的查询描述符或描述模式数据进行匹配, 修改符合查询条件的媒体数据, 将修改后的媒体数据保存在一个新的媒体数据库中。在该应用中, 查询是可选的, 如果没有查询条件就将媒体数据库中所有媒体都进行修改, 生成一个新的媒体数据库。

其过程如图4所示:

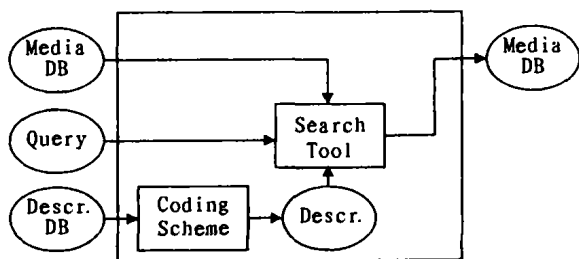


图4

4. 描述过滤应用 将描述数据库文件按照相应的编码模式装载到内存中, 根据输入的查询条件进行搜索, 将符合查询条件的描述符或描述模式信息依照相应的编码模式保存到另一个描述数据库中。

其过程如图5所示:

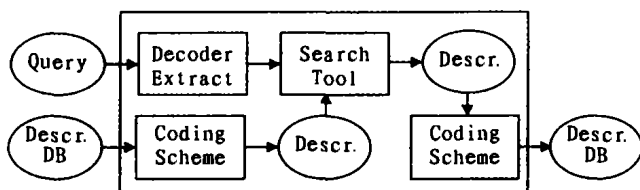


图5

以上四种应用组成了 MPEG-7的“关键应用”, 也就是基本的应用类型。现实中的应用或者就是四种应用的一种, 或者是由这四种应用组合而成的复杂应用。

2.2 XM 参考软件模块

MPEG-7 XM 参考软件包括了以下一些模块:

1. 媒体数据 要处理的各种多媒体数据在内存中由 MultiMedia class 来表示。对于视频流并不装载到内存中, 而

是只装载解码后视频帧。

2. 媒体解码器 支持各种可能的输入媒体格式:

- 以 WAV 文件存储的音频数据 由 Afsp 库来处理(不包含在 XM 软件中)。
- MPEG-1, MPEG-2 视频流 由 MPEG-1, MPEG-2 编码器和解码器来处理。
- MPEG-1 视频 motion vector 由 MPEG-1 video motion vector extractor(MPEG-1 视频动作向量抽取器)来处理。
- 各种静态图像(JPEG, GIF, PNM 等等) 由 ImageMagick 库来处理(不包含在 XM 软件中)。
- 3D Objects 读取 3D 对象, 创建它们的 3D 形状描述器。
- Key Point 关键点 从文件中读取关键点列表。

在处理视频流时, 先通过解码器将视频流解码成单个的帧, 将每个帧看成是一幅静态图像装载到内存中进行处理。

3. 抽取工具 完成各种多媒体数据的各种特征的抽取工作。抽取工具在实现中不是标准化的部分。抽取工具的输入是被抽取的媒体数据, 并存储相应抽取的描述信息到描述符。

在处理视频流时, 不可能同时提供所有的数据, 抽取工作是基于帧进行的, 完成这种抽取需要三个函数(步骤):

- InitExtracting 初始化抽取操作, 在处理第一帧之前调用。
- StartExtracting 完成抽取处理操作, 处理每一帧。
- PostExtracting 完成抽取事后工作, 在所有帧都处理完后调用。

4. 描述符和描述模式类 描述符类和描述模式类存储媒体描述数据。XM 软件中每个 D 或 DS 类都是代表 MPEG-7 标准的标准化部分。

XM 软件实现采用两种方式来实现 D 或 DS 类, 所有的视频描述符都单独采用一个 C++ 类来实现, 所有其他的描述符和描述模式采用一个通用的 GenericDS 类来实现。GenericDS 类并不专门实现一种数据结构, 它是 XM 软件与 XML 剖析器的接口。当前使用的是具有 DOM-API(Data Object Model - Application Programming Interface) 的 XML 剖析器。

5. 编码模式 包括 D 或 DS 的标准化的编码、解码器。在绝大部分情况下, 编码模式是由 DDL 模式定义来定义的。在这种情况下, 编码就是将描述信息写到文件中, 解码是剖析描述文件并将描述信息载入内存。

如果单个编码模式存在, 它就是 MPEG-7 标准的一个标准化部分。视觉描述符可以使用编码模式来将一个描述解码、编码成二进制形式。对于通过 GenericDS 类来实现的 Ds 和 DSs 不存在编码模式。对于这些 Ds 和 DSs 使用 GenericDSCS 类实现了一个编码模式盒子(coding scheme box), GenericDSCS 类是 XML 剖析器库的一个读写接口。

MPEG-7 标准化了 ASCII 表示的 XML 文件和描述信息的二进制表示(BiM)。

6. 匹配工具 和抽取工具一样, 匹配工具也不是 MPEG-7 标准的标准化部分。匹配工具是为每个描述符和描述模式定义的特定的搜索或匹配方法。

7. 应用 在 MPEG-7 软件框架中, 应用与一个 D 或 DS 相关。在 MPEG-7 中存在大量的 D 或 DS, 因此也就存在大量的应用。那些从原始媒体数据生成描述符和描述模式数据库的应用, 叫做抽取应用, 也叫做服务器应用(Server Applications), 使用生成的描述符或描述模式数据库的应用叫做客户

端应用(Client Applications)。

对于低级别的 D 或 DS(如视频描述符),抽取应用是必须的。这些描述信息可以自动地从媒体数据中抽取出来。对于那些高级别的,不能够自动抽取的 D 或 DS,抽取过程除了媒体内容本身外还需要其他的一些信息。

8. 其他一些辅助模块 提供了一些基本工具和设施,方便其他模块的实现。

2.3 XM 参考软件描述符和描述模式的实现

2.3.1 标准化描述符和描述模式的实现 XM 参考软件实现了每个标准化的描述符。XM 为每个描述符的实现提供了统一的格式。可以方便地添加新的描述符,同时也方便了对于现有描述符的修改。每个描述符的实现包括五个部分:

1)描述符自身的实现

XM 软件中每个低级别的描述符都对应着一个 C++ 类文件,它们都继承一个公共的描述符类:Descriptor,并通过声明一个描述符接口类为友元关系(friend)将内部具体的操作屏蔽起来。该描述符接口类继承一个描述符接口 ABC 类,该 ABC 类继承一个公共的 ABC 类:I-InterfaceABC 类。类流图如图6所示:

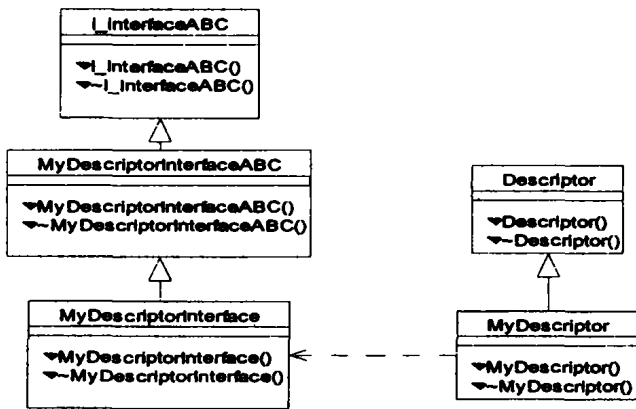


图6

2)描述符编码模式的实现

每个描述符编码模式也对应着一个 C++ 类文件,都继承一个公共的编码模式类:DescriptionCodingEngine,并通过声明一个描述符编码模式接口类为友元关系将内部具体的操作屏蔽起来。该描述符编码模式接口类继承一个描述符编码模式接口 ABC 类,该 ABC 类继承一个公共的 ABC 类:I-InterfaceABC 类。类流图如图7所示:

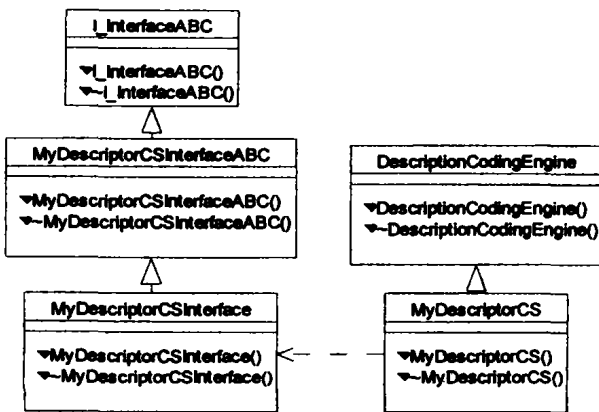


图7

3)描述符抽取工具(器)的实现

每个描述符抽取工具也对应着一个 C++ 类文件,都继承一个公共的抽取器类:DescriptorExtractor,并通过声明一个描述符抽取器接口类为友元关系将内部具体的操作屏蔽起来。该描述符抽取器接口类继承一个描述符抽取器接口 ABC 类,该 ABC 类继承一个公共的 ABC 类:I-InterfaceABC 类。类流图如图8所示:

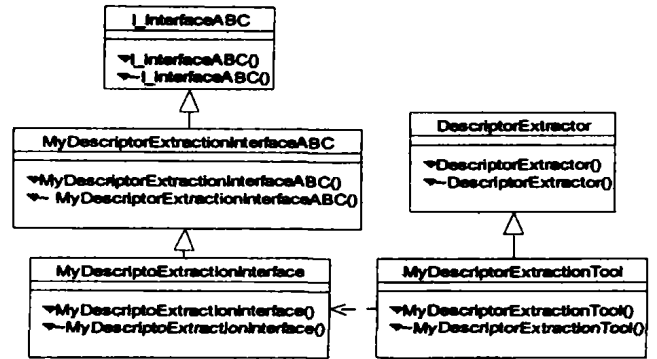


图8

4)描述符搜索工具(器)的实现

每个描述符搜索工具也对应着一个 C++ 类文件,都继承一个公共的搜索类:Search,并通过声明一个描述符搜索接口类为友元关系将内部具体的操作屏蔽起来。该描述符搜索接口类继承一个描述符搜索接口 ABC 类,该 ABC 类继承一个公共的 ABC 类:I-SearchInterfaceABC 类。类流图如图9所示:

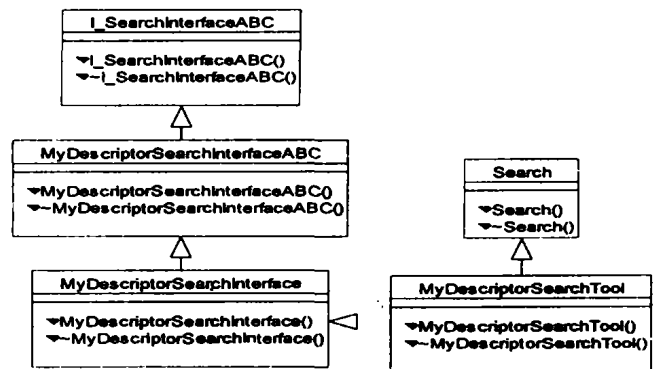


图9

5)描述符服务器应用(ServerApplication)和客户端应用(ClientApplication)的实现

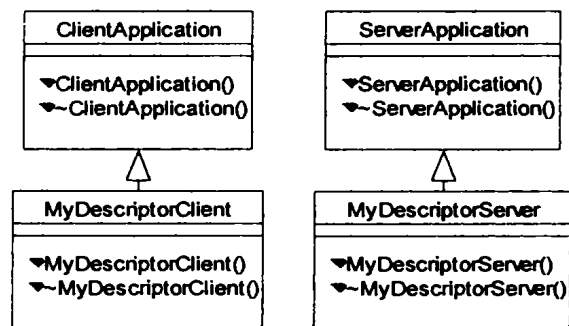


图10

每个描述符的服务器应用对应着一个 C++ 类文件,服务器应用类继承一个公共的服务器应用类:ServerApplication。每个描述符的客户端应用也对应着一个 C++ 类文件,客户端应用类继承一个公共的客户端应用类:ClientApplication。类流图如图10所示。

2.3.2 其他描述符和描述模式的实现 对于非标准化的描述符和描述模式,MPEG-7采用一个 GenericDS 类来实现。GenericDS 类没有专门实现一种数据结构,它封装了 XM 软件中有关 XML 剖析器的各种操作。GenericDS 类继承了一个公共的描述符模式类:DescriptionScheme。并通过一个 GenericDSInterface 类将内部具体的操作屏蔽起来。该 GenericDSInterface 类继承 GenericDSInterfaceABC 类,该 ABC 类继承一个公共的 ABC 类:I_InterfaceABC 类。类流图如图11所示:

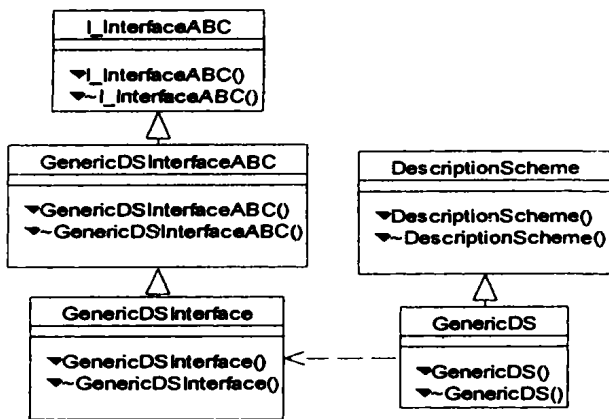


图11

通过 GenericDS 实现的描述符和描述模式不存在单独的编码模式类,通过使用的一个一般性的编码模式类:GenericDSCS 类来实现编码、解码。GenericDSCS 类实现了 XML 剖析器的一个读写接口。GenericDSCS 类继承一个公共的编码模式类:DescriptionCodingEngine。并通过声明 GenericDSCSInterface 类为友元关系将内部具体的操作屏蔽起来。该 GenericDSCSInterface 类继承了 GenericDSCSInterfaceABC 类,该 GenericDSCSInterfaceABC 类继承一个公共的 ABC 类:I_InterfaceABC 类。类流图如图12所示:

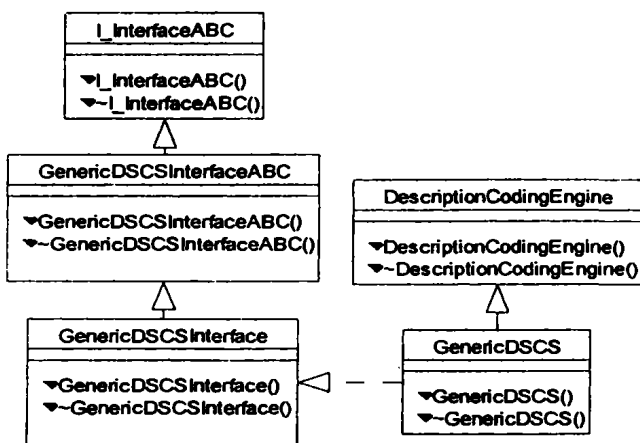


图12

通过 GenericDS 实现的 D 或 DS 不是 MPEG-7 标准标准化的一部分,不存在相应的抽取工具和匹配工具,以及相应的

应用。随着 MPEG-7 的不断发展,一些由 GenericDS 实现的 Ds 或 DSs 有可能成为标准化的部分,在 MPEG-7 的 XM 软件将实现它们的抽取工具类、匹配工具类以及服务器和客户端应用。

3 MPEG-7 在视频音像资源事务管理系统 (AVTMS) 中的潜在应用

AVTMS 管理一台或多台 JVC 的 DVD 光盘库系统 (DVD 光盘库系统管理一个或多个光盘库),为用户提供一个操作 DVD 光盘库的图形接口。主要功能模块有:

- 著录子系统,实现视频音像资料入库功能
- 检索子系统,实现对入库视频音像资料检索功能
- 统计子系统,统计用户对入库视频音像资料的访问统计
- 导入导出子系统,著录信息的导入与导出,便于视频音像资料的交流
- 自动播放子系统,自动播放用户检索的视频音像资料
- 系统管理和用户管理子系统,管理访问本系统的用户和本系统的日常维护操作

AVTMS 管理的音像资源有 CD, VCD, DVD 等多种不同格式。目前采用 MARC (Machine Readable Catalogue or Cataloguing, 简称 MARC) 格式对音像资源进行统一编码,用户通过 IE 等浏览器检索光盘库系统中存储的音像资源。AVTMS 目前已有了好几家安装单位,从目前应用的情况看,采用 MARC 格式编码音像资源信息存在以下几个方面的问题:

1. MARC 是非结构化可变长的形式,因此它在目前的关系型数据库中的存储和检索有一定的困难。
2. MARC 记录的生成需要有专门知识,在使用过程中往往也需要一些专门知识。
3. MARC 格式的一种数据交换格式,可以和同样采用 MARC 格式的系统交换多媒体信息,但是无法和采用其他格式的系统进行媒体信息的交换。

这些问题的存在使得 AVTMS 只能是一种专业的音像资源管理系统:管理特定的音像信息,由专业人员使用,采用专门的搜索系统。但是 MPEG-7 的出现使得 AVTMS 可以摆脱这些限制。MPEG-7 可以统一处理我们生活环境中的所有多媒体信息,包括多媒体信息的表示和多种媒体相互之间的关系。其通用的、标准化的媒体表示形式方便了不同系统之间媒体资源交换。采用 MPEG-7 的搜索引擎支持基于内容的检索或者 query-by-example (按示例搜索) 的搜索方式使得用户可以更直观地搜索到他/她所需要的媒体,而无须任何专门的知识。

因此可以为 AVTMS 系统做如下改进:

1. 保留原有的 MARC 格式的著录方式,另外添加基于 MPEG-7 标准的著录方法,生成音像资源的 MPEG-7 描述信息。
2. 保留原有的基于文本的检索方式,增加基于内容/query-by-example 的检索方法。

结束语 MPEG-7 已经成为正式的国际标准。它的发展阶段包括:1998年10月开始征求标准;1999年2月进入评估阶段;1999年12月第一版草稿推出;2000年10月推出委员会审查草稿;2001年2月推出最终版委员会会议草稿;2001年7月推出国际标准初稿;2001年9月国际标准定案。MPEG-7 标准的制定将使得基于内容的搜索引擎出现并最终取代传统的基于文本

的搜索引擎成为可能,从而可以方便人们对多媒体数据资源的访问和提高多媒体数据资源的利用率。

附录

系统参考软件(Systems reference software)

Bit Stream Encoder/Decoder (to be submitted)

Access Unit Navigation (to be submitted)

DDL Parser External XML parser supporting the DOM API (e. g. , “xerces” XML parser from Apache)

DDL Validation parser (to be submitted)

视频参考软件(Video Reference software)

种描述符的抽取、存储、匹配等功能。每个描述符对应一个

视频参考软件实现描述视频信息各种描述符,包括各 C++类文件。

描述符名字	参考软件文件名或类名	功能说明
Grid Layout	GridLayout	将一幅图像分成几个长方形区域,每个区域都可以用其他的描述符如 Color, Texture 进行描述。
Time Series	TemporalSeries	描述一个视频片段的基于时间的描述符系列,以提供 image 与 video-frame 的匹配和 video-frame 与 video-frame 的匹配功能。分为 RegularTimeSeries 和 IrregularTimeSeries 两种。
Multiple View	3DshapeSpectrum	将一个3D物体的不同角度的2D描述符组合起来,形成这个物体的3D视图。
Spatial 2D Coordinates	Spatial2Dcoordinates	定义其他相关 Ds/DSs 参考使用的2D空间坐标系。支持两种类型的坐标系:“局部的”和“综合的”。
Temporal Interpolation	TemporalInterporation	描述一个使用连接多项式的时间的插入修改,用来描述一个随着时间改变的多维可变量。
Color Space	ColorSpace	定义其他基于颜色的描述要使用的颜色空间。目前支持: ◇R,G,B ◇Y,Cr,Cb ◇H,S,V ◇HMMD ◇Linear transformation matrix with reference to R,G,B ◇Monochrome
Color Quantization	ColorQuant	定义一个颜色空间的量化。支持线性和非线性的量化器。用于为其他的颜色描述符描述一个统一的颜色量化。
Dominant Color	DominantColor	描述任意形状区域的主颜色集合。
Scalable Color	ScalableColor	描述一个 HSV 颜色空间内的颜色柱状体,通过 Haar 变换来编码。它的二进制表示在很大的数据速率范围内在 bin 的数目和位的表示精度上都是可扩展的。
Color Layout	ColorLayout	描述一个空间上的颜色分布状态。支持 image-to-image 的匹配和 video-clip to video-clip 匹配。
Color Structure	ColorStructure	定义一幅图像的颜色结构。主要支持 image-to-image 的匹配。
GoF/GoP Color	GoFColorHist	Group-Of-Frames/Pictures(GoF/GoP) 颜色描述符用于描述一系列的图像或视频帧的颜色特征。
Homogeneous Texture	HomoTexture	描述 image-to-image 匹配中质地的相似性。
Texture Browsing	TextureBrowsing	详细说明了一个质地浏览描述符。它描述质地的感觉上的特征如规则性,粗糙和方向性。
Edge Histogram	EdgeHist	描述5种类型边的空间上的分布,这5种边是四种方向性的边,一个水平的,一个垂直的,两个对角的,和一个非方向性的边。
Region Shape	RegionShape	描述任意区域的形状。
Contour Shape	ContourShape	描述一个2D物体,或者一幅图像或者视频系列的一个区域的封闭轮廓。
Shape 3D	Shape3D	为一个3D网状模型提供一个内在的形态描述。
Camera Motion	CameraMotion	描述3D照相机运动参数。它支持 fixed, panning, tracking, tilting, booming, zooming, dollying 和 rolling 等照相机操作。
Motion Trajectory	ObjectMotionTraj	定义对象的运动轨迹。
Parametric Motion	ParametricObjectMotion	
Motion Activity	MotionActivity	描述一个活动的属性,如动作的强度、速度、方向等。
Region Locator	RegionLocator	定义一个区域的位置。
Spatio-Temporal Locator	SpatioTemporalLocator	描述视频系列中时间空间上的一个区域。
Face Recognition	FaceRecognition	代表一个面部向量到4个特征向量组的投影。

音频参考软件(Audio reference software)

C++类文件。目前大部分音频信息的描述符尚未最终确定下来。

视频参考软件实现描述视频信息各种描述符,包括各种描述符的抽取、存储、匹配等功能。每个描述符对应一个

描述符名字	参考软件文件名或类名	说明
Spoken Content	SpokenContent	描述音频流中说的词语
Harmonic Instrument Timbre	HarmonicInstrumentTimbre	描述管弦乐器的音质
Percussive Instrument Timbre	PercussiveInstrumentTimbre	描述打击乐器的音质
Silence	(to be submitted)	
Audio Spectrum Basis	(to be submitted)	
Sound Effect State Path	(to be submitted)	
Sound Effect Classifier	(to be submitted)	
Sound Effect Model Type	(to be submitted)	
Melody Contour	(to be submitted)	
Melody	(to be submitted)	

多媒体描述模式参考软件 (Multimedia Description Scheme Reference Software)

描述符名字	参考软件文件名或类名	说明
Segment DS	Segment	提供了派生出专门的段描述模式的一个抽象类型。
Sequential Summary DS	SequentialSummary	描述一个单一的 AV 摘要, 这个摘要可能包括一系列的图像或者视频帧。
Meta Media DS	MediaMeta	
VariationSet DS	Variation	描述多媒体程序的不同变种。
Vector DS	VectorX	
Model DS	Model	详细说明了形成刻画多媒体内容属性和特征的不同专门模型的基础的一个抽象类型。
Time data type	Time	指定一个时间点或时间间隔的描述模式, 它使用以下几个描述符来表示时间规范: <ul style="list-style-type: none"> ◇ TimePoint 指定一个时间点的元素 ◇ RelTimePoint 指定一个相对时间的元素 ◇ RelIncrTimePoint 指定一个相对增加的时间点的元素 ◇ Duration 指定一段持续时间的元素 ◇ IncrDuration 指定以时间单位计数的一段时间的元素
Fine State Model DS (state transition data type)	StateTransitionModel	通过给定各状态, 状态转换的可能性和初始状态可能性来描述一个状态转换模型。
Mosaic DS	Mosaic	描述一个视频段的全景或马赛克视图。
Collection Structure DS	CollectionStructure	描述多个集合, 集合模型和簇模型之间的联合关系。
Space Frequency View DS	SpaceFrequencyView	描述在多维空间和频率上相似的一幅图像, 视频或音频信号的视图。
Video Text DS	VideoText	描述与文本或标题相关的视频数据的时间空间区域。
User Preference DS	UserPreference	描述用户消费多媒体内容的偏好。
Transcoding Hints D (importance attribute)	ImportanceHint	描述有关媒体语义内容的重要性。
Media Transcoding Hints D (motion hint data type)	Motion Hint	描述代码转换器的动作线索。
Media Transcoding Hints D (difficulty attribute)	Difficulty Hint	描述一个媒体段中转换媒体的困难级别。
Classification DS (Media Review data type)	MediaReview	描述对多媒体内容的一个媒体评论。
ClassificationPreference DS	MediaReviewPreference?	描述与媒体分类描述的相关的用户偏好。
Affective DS	AffectiveSegment	描述观众对多媒体内容的情感反应。
Matching Hint D	MatchingHint	描述音视频描述工具的实例或者各部分的相对重要性。
Point Of View D	PointOfView	描述在给定一个具体观点时一段媒体的相对重要性。
Sound Property D	SoundProperty	描述与一个音频片段或一个音频幻灯片放映组件相关联的某些属性。
Text Property D	Summary (Text Property)	描述与 AV 摘要相关联的原文信息的某些属性。
Entity-Relation Graph DS	SegmentRelGraph	
Dependency Structure data type	DependStructure	描述组成一个句子的语法元素之间的语义依赖性结构一个文本注解的结构。
StructureIndex	StructureIndex	
Media Quality D	MediaQuality	描述 AV 内容的质量评价信息。
Usage Preferences DS	Usage Preferences DS	描述关于过滤, 搜索和浏览多媒体内容的用户偏好。
Classification DS (parental guidance data type)	ParentalGuidance	描述多媒体内容的分类。
Media Format D	MediaFormat	描述关于文件格式和 MediaProfile 编码参数的信息。
Ordering Key DS	OrderingKey	描述相同类型描述排序的线索。

(下转第81页)

用于改正 HTML 文档中的常见错误并生成格式编排良好的等价文档,还可以使用 Tidy 来生成 XHTML^[16](XML 的子集)格式的文档。

(2) 找数据中的引用点。无论是在 Web 页面还是 XHTML 视图中的绝大多数信息都与我们完全无关。在这一步中我们的任务就是在 XML 树中找出一个特定区域,并从中抽取我们感兴趣的数据而无需关心外来信息。完成这一任务的最简单的办法通常是,首先检查 Web 页面,只需要看一下页面就可以知道信息位于页面的位置(称之为锚),然后使用 XSL 来转换我们的 XML,利用 Xpath^[17]表达式来指定从根元素到锚的路径。如:/html/body/center/table [6]/tr [2]/td [2]/table [2]/tr/td/table [6],但是这个方法的缺陷就在于会导致我们对页面布局的修改非常容易遭到破坏。较好的方法是根据周围的内容指定锚。通过这个方法,我们对上例中的 Xpath 表达式重新构造://table[starts-with(tr/td/font/b,'Appear Temp')],这样我们就很容易找到表格中粗体显示的 'Appear Temp' 信息了。

(3) 将数据映射成 XML。从前一步中得到锚,我们可以创建实际抽取数据的代码。这个代码将以 XSL 文件的形式出现。XSL 文件的目的是标识锚,指定如何从这个位置获取我们正在查找的数据,并且用我们所需的格式构造一个 XML 输出文件。

小结 Web 信息抽取是一个年轻的研究领域,尽管目前该领域研究已经取得了一定的进展,但仍然存在一些问题。首先,信息抽取系统的准确性和健壮性有待提高;其次,在一个新领域上建立信息抽取系统需要许多该领域专家和熟悉 NLP 系统的计算语言学家的共同努力,既费时又费力。随着网络在国内的迅猛发展,Web 信息抽取会变得越来越重要,希望有更多更好的技术能够应用到该领域。

参考文献

- 1 Cardie C. Empirical methods in information extraction. AI Magazine, 1997, 18(4): 65~79
- 2 朱靖波,姚天顺. 中文信息自动抽取. 东北大学学报, 1998, 19(1)
- 3 Muslea I. Extraction Patterns for Information Extraction tasks: A survey. In: AAAI-99 Workshop on Machine Learning for Infor-

- mation Extraction, 1999
- 4 朱明. 互联网信息智能搜索和获取方法研究: [中国科学技术大学博士论文]. 2001
- 5 Soderland S. Learning to extract text-based information from the world wide web. In: Proc. of Third Intl. Conf. on Knowledge Discovery and Data Mining (KDD-97)
- 6 Huffman S B. Learning information extraction patterns from examples. In Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing, Springer Verlag, Berlin, 1996, 1040: 246~260
- 7 Kim J, Moldovan D. Acquisition of Semantic Patterns for Information Extraction from Corpora. In: Proc. of the Ninth IEEE Conf. on Artificial Intelligence for Applications, Los Alamitos, CA: IEEE Computer Society Press, 1993. 171~176
- 8 Califf M, Mooney R. Relational Learning of Pattern-Match Rules for Information Extraction. Working Papers of the ACL 97 Workshop on Natural Language Learning, 1997. 9~15
- 9 Soderland S. Learning information extraction rules for semi-structured and free text. Machine Learning, 1999, 34(1-3): 233~272
- 10 Freitag D. Information extraction from html: Application of a general learning approach. In: Proc. of the 15th Conf. on Artificial Intelligence (AAAI-98), 1998. 517~523
- 11 Muslea I, Minton S, Knoblock C. A hierarchical approach to wrapper induction. In: Proc. of the Third Intl. Conf. on Autonomous Agents(AA-99)
- 12 Guan T, Wong KF. KPS: a Web Information Mining Algorithm. Computer Networks, Elsevier, 1999, 31: 1495~1507
- 13 Myllymaki J. Effective Web data extraction with standard xml technologies. WWW10, Hong Kong ACM 1-58113-348-0/01/0005, 2001
- 14 Extensible Markup Language (XML). W3C Recommendation, Feb. 1998. <http://www.w3.org/TR/REC-xml>
- 15 HTML Tidy. <http://www.w3.org/People/Raggett/tidy/>
- 16 XHTML: The Extensible HyperText Markup Language. W3C Recommendation, January 2000. <http://www.w3.org/TR/xhtml1>
- 17 XML Path Language (XPath). W3C Recommendation, November 1999. <http://www.w3.org/TR/xpath.html>

(上接第37页)

Semantic DS	Semantic	描述了由多媒体内容模板所描绘或者与之有关的一个叙述性的世界。
Package DS	Package	描述了工具的一种树状组织。
Video Editing DS	(To be submitted)	
Summary Preferences DS	SummaryPreference	详细说明用户对多媒体内容的非线性导航和访问偏好。
Extended Textual Type	ExtentedTextualType	
Phonetic D	Phonetic	描述各种语音信息。

参考文献

- 1 MPEG-7标准文档 Overview of the Mpeg-7 Standard N4031
- 2 MPEG-7标准文档 MPEG-7 Context and Objectives N2861
- 3 MPEG-7标准文档 MPEG-7 Requirements Document N4320
- 4 MPEG-7标准文档 MPEG-7 Applications Document N3934

- 5 MPEG-7标准文档 MPEG-7 Projects and Demos N4034
- 6 MPEG-7标准文档 MPEG-7 Interoperability, Conformance Testing and Profiling Version 2 N4039
- 7 <http://archive.dstc.edu.au/mpeg7-ddl/> MPEG-7 DDL 主页
- 8 MPEG-7 eXperimentation Model (XM) Software version 5. 5